Title	Some solution of a cooperative m-person doscounted Markov game(Studies on Control,Learning and Their Related Topics)
Author(s)	田中, 謙輔
Citation	数理解析研究所講究録 (1985), 557: 32-45
Issue Date	1985-04
URL	http://hdl.handle.net/2433/98987
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録
557巻 1985年 32-45

Some solution of a cooperative m-person discounted Markov game

新潟大・理　田中謙輔（Tanaka Kensuke）

§1. 問題の定式化について

ここでは割引因子をもつ協力m人マルコフ・ゲームを次のような（2m＋3）個の組み

$$(S, A^1, A^2, \cdots, A^m, q, r^1, r^2, \cdots r^m, \beta) \qquad (1)$$

で与える．ただし

(1)　$S \equiv \{1, 2, 3, \cdots\}$ はゲームの状態空間

(2)　A^i, $i=1,2,\cdots,m$, はiプレイヤーの行動空間と呼ばれ，コンパクトな距離空間

(3)　q は $\bar{a} \equiv (a^1, a^2, \cdots, a^m) \in \prod_{i=1}^{m} A^i \equiv A$ に対応するSの上の推移確率測度 $q(\cdot \mid s, \bar{a})$, $s \in S$

(4)　r^i, $i=1,2,\cdots,m$, はiプレイヤーの損失関数と呼ばれ$S \times A$の上で定義されている実数値関数

(5)　βは割引因子，$0 < \beta < 1$．

このゲームシステムでは，すべてのプレイヤーが各時

点 $t=1,2,\cdots$ でゲームプロセスの状態を観測し，各人が協力して現時点の状態 $s\in S$ のみに関係して確率的に行動 $a^i\in A^i$ を選択する．この結果 i プレイヤーは損失 $r^i(s,\bar{a})$，$\bar{a}=(a_1,a_2,\cdots,a_m)\in A$ を受ける．この後ゲームプロセスは推移確率 $q(s'|s,\bar{a})$ に従って次の新しい状態 $s'\in S$ に移りシステムはくり返される．このとき各時点で，ある別払率（rate of transfer for side-payment）$d=(d_1,d_2,\cdots,d_m)$ に関する全体的損失関数

$$\sum_{i=1}^{m} d_i r^i(s,\bar{a})$$

を最小とする $\bar{a}\in A$ を選択するように協力して，各プレイヤーは与えられている凸錐 D の支配構造のもとで最適な解を得るようにする．

この話しを通して用いられる多重戦略は $\pi=(\pi^1,\pi^2,\cdots\pi^m)$ で表わす．ここで，各 π^i は i プレイヤーの戦略で各時点 t の状態 S_t に対応する $(A^i,\beta(A^i))$ の上の確率測度 $\pi^i_t(\cdot|S_t)$ の列によって与えられる，ただし $\beta(A^i)$ は行動空間 A^i 上の Borel field である．特に π^i_t が time t に無関係，i.e., $\pi^i_t=\mu^i\in[P(A^i)]^S$ ならば π^i は定常戦略と呼ばれ，π^i は μ^i と同一視できる，ただし $P(A^i)$ は測度空間 $(A^i,\beta(A^i))$ 上の確率測度の全体である．このような各プレイヤーの戦略の

全体をΠ^iと表わし，多重戦略の全体は$\Pi \equiv \prod_{i=1}^{m} \Pi^i$と書くことにする．

このとき初期状態$s \in S$と多重戦略$\bar{\pi} = (\pi^1, \pi^2, \cdots, \pi^m)$に対して　iプレイヤーの総期待損失は

$$I^i(\bar{\pi})(s) = \sum_{t=1}^{\infty} \beta^{t-1} E_{\bar{\pi}} [r^i(s_t, t, \bar{\pi}) \mid s_1 = s]$$

で与えられる．ここで

$$\pi^i = (\pi^i_1, \pi^i_2, \cdots, \pi^i_t, \cdots\cdots\quad)$$

$$r^i(s_t, t, \bar{\pi}) = \int_A \cdots \int r^i(s_t, \bar{a}) d\bar{\pi}_t(\bar{a} \mid s_t)$$

$$d\bar{\pi}_t(\bar{a} \mid s_t) = \prod_{i=1}^{m} d\pi^i_t(a^i \mid s_t)$$

$$\bar{a} = (a^1, \cdots, a^m) \in \prod_{i=1}^{m} A^i \equiv A .$$

また，各$I^i(\bar{\pi})(s)$，$i = 1, 2, \cdots, m$，のベクトル表示を次のようにする．

$$I(\bar{\pi})(s) = (I^1(\bar{\pi})(s), I^2(\bar{\pi})(s), \cdots, I^m(\bar{\pi})(s))$$

$$= \sum_{t=1}^{\infty} \beta^{t-1} E_{\bar{\pi}} [r(s_t, t, \bar{\pi}) \mid s_1 = s],$$

ただし

$$r(s_t, t, \bar{\pi}) = (r^1(s_t, t, \bar{\pi}), r^2(s_t, t, \bar{\pi}), \cdots, r^m(s_t, t, \bar{\pi}))$$

で

$$E_{\bar{\pi}}[r(s_t, t, \bar{\pi}) \mid s_1 = s] = \left(\cdots\cdots, E_{\bar{\pi}}[r^i(s_t, t, \bar{\pi}) \mid s_1 = s], \cdots\cdots \right)_{i=1}^{m}.$$

§2. 準備と記号

$R^m \supset E$ に対して，次のような記号を導入する．clE は E の閉包，$intE$ は E の内部，$E^* \equiv \{ y \in R^m \mid \langle x, y \rangle \geqq 0, \forall x \in E \}$ は E の正極錐，$[E] \equiv \{ y \in R^m \mid y = \lambda x, x \in E, \lambda \in R_+ \}$ は E によって作られる錐とする．

ここの話しを通して，次のような条件をみたす部分集合 $L \subset R^m$ を与える；

(i) $L \not\ni 0$

(ii) $L \ni e = (1, 1, \cdots, 1)$

(iii) $L \cup \{0\} \equiv D$ は原点を頂点とする凸錐．

さらに次のような集合を導入する；

$$L^+ \equiv \{ y \in R^m \mid \langle x, y \rangle > 0, \forall x \in L \}$$

と

$$L_1^+ \equiv \{ y \in R^m \mid y \in L^+, |y| = \sum_{i=1}^{m} |y_i| = 1 \}.$$

ここで集合 L_1^+ は通常の非負の要素をもっている割払率の概念を拡張していることになり，$L_1^+ \ni d = (d_1, d_2, \cdots, d_m)$ を割払率の集合とみることにする．

<u>定義 1.</u>　多重戦略 $\bar{\pi}^* = (\pi^{*1}, \pi^{*2}, \cdots, \pi^{*m})$ に対して

$$I(\bar{\pi}^*)(s) \in I(\bar{\pi})(s) + L$$

となる $\bar{\pi} \in \Pi$ が存在しないとき，$\bar{\pi}^*$ は初期状態 s に対する D-solution と呼ぶ．ただし $\Pi = \prod_{i=1}^{m} \Pi^i$ で各 Π^i は i プレイヤーの戦略の全体である．

<u>注意.</u>　一般には，閉凸集合 E に対して $L \equiv \mathrm{int}\, E$ 又は $L \equiv E - \{0\}$ であるとき，定義 1 で与えられる $\bar{\pi}^*$ は E-weak solution 又は E-strong solution と呼ばれている．

ここで $E_s \equiv \{ I(\bar{\pi})(s) \mid {}^\forall \bar{\pi} \in \Pi \}$ を用いて，$Ext[E_s|D]$ を初期状態 s に対する総期待多重損失の全ての D-solution に対応する集合を表わすことにする．

<u>補助定理 1.</u>　$d \in L_1^+$,

$$\langle d, I(\bar{\pi}_d)(s) \rangle = \min_{\bar{\pi}} \langle d, I(\bar{\pi})(s) \rangle \qquad (2)$$

とする．このとき $\bar{\pi}_d$ は初期状態 $s \in S$ に対応する D-solution となる．

証明　$\bar{\pi}_d$ が D-solution でないと仮定すると，

$$I(\bar{\pi}_d)(s) \in I(\bar{\pi})(s) + L$$

をみたす $\bar{\pi} \in \Pi$ が存在する, i.e.,

$$I(\bar{\pi}_d)(s) = I(\bar{\pi})(s) + \hat{\alpha}$$

をみたす $\hat{\alpha} \in L$ が存在する．よって $L^+ \ni d$ との内積を作ることで

$$\langle d, I(\bar{\pi})(s) \rangle < \langle d, I(\bar{\pi}_d)(s) \rangle$$

となり（2）に反するのでこの補助定理は成立する．

§3. 割引因子をもつマルコフ・ゲームにおける D-solution の存在について

定常多重戦略 $\bar{\mu} = (\mu^1, \mu^2, \cdots, \mu^m) \in \prod_{i=1}^{m} P(A^i) \equiv P(A)$ が用いられたとき，各プレイヤーの損失を

$$r^i(s, \bar{\mu}) \equiv \int \cdots \int_A r^i(s, \bar{a})\, d\bar{\mu}(\bar{a}) \;,$$

多重損失を

$$r(s, \bar{\mu}) = (r^1(s, \bar{\mu}), r^2(s, \bar{\mu}), \cdots, r^m(s, \bar{\mu}))$$

と書き，推移確率測度は

$$q(\cdot \mid s, \bar{\mu}) \equiv \int \cdots \int_A q(\cdot \mid s, \bar{a})\, d\bar{\mu}(\bar{a})$$

で与えられる，ただし

$$\bar{a} = (a^1, a^2, \cdots, a^m) \in A = \prod_{i=1}^{m} A^i, \quad d\bar{\mu}(\bar{a}) = \prod_{i=1}^{m} d\mu^i(a^i) \;.$$

$C(A^i)$ を A^i の上の実数値連続関数の全体とすると，A^i がコンパクトな距離空間であるから supnorm の導入によって $C(A^i)$ は可分なバナッハ空間となる．$C(A^i)^* = M(A^i, \beta(A^i))$ は有界な正則測度空間（bounded regular measure space）となり $P(A^i)$ は $C(A^i)^*$ の中で汎弱コンパクト（weak*compact）である．よって $P(A) \equiv \prod_{i=1}^{m} P(A^i)$ は $C(A)^*$（$C(A) \equiv \prod_{i=1}^{m} C(A^i)$）の中で可分で汎弱コンパクトとなる位相が導入されている．

今 $C^b(S)$ を S 上の有界な（連続）実数値関数の全体とし，ゲームシステムが D-solution を持つように q と r^i の上に次のような追加条件を与える．

(A1)　$(s', s) \in S \times S$ に対して $q(s' | s, \bar{a})$ は $\bar{a} \in A$ に関して連続である．

(A2)　i プレイヤーの損失関数 $r^i(s, \bar{a})$ は $S \times A$ の上で有界で，各状態 $s \in S$ に対して A の上で連続である．

ここで $L_1^+ \ni d$ に対して $T_d : C^b(S) \to C^b(S)$ を次のように定義する

$$T_d u(s) \equiv \min_{\bar{\mu} \in P(A)} \left[\langle d, r(s, \bar{\mu}) \rangle + \beta \sum_{s'} u(s') q(s' | s, \bar{\mu}) \right]. \qquad (3)$$

さらに簡単化のために

$$L_d(\bar{\mu})u(s) \equiv \langle d, r(s,\bar{\mu})\rangle + \beta\sum_{s'} u(s')\,q(s'|s,\bar{\mu}) \qquad (4)$$

を導入して

$$T_d u(s) = \min_{\mu} L_d(\bar{\mu})u(s)$$

と書くことにする．このとき T_d は $0<\beta<1$ によって 縮小写像となるので次の定理が得られる．

<u>定理 1.</u>　（A1）と（A2）の仮定のもとで，$L_1^+ \ni d$ に対して

$$\langle d, I(\bar{\mu}_d^*)(s)\rangle \leqq \langle d, I(\pi)(s)\rangle \quad {}^\forall\pi\in\Pi \qquad (5)$$

をみたす定常多重戦略 $\bar{\mu}_d^*$ が存在し，この戦略は初期状態 s に対する D-solution となっている, i.e.,

$$I(\bar{\mu}_d^*)(s) \in \mathrm{Ext}[E_s|D].$$

証明　（3）で定義されている写像 $T_d : C^b(S) \to C^b(S)$ は 割引因子 $0<\beta<1$ より縮小写像となっている．よって T_d の不動点 $u^*\in C^b(S)$ が存在する, i.e.,

$$u^*(s) = T_d u^*(s) = \min_{\mu} L_d(\bar{\mu})u^*(s). \qquad (6)$$

さらに，$L_d(\bar{\mu})u^*(s)$ はコンパクト集合 $P(A)$ の上で連続であるから（6）より,

$$u^*(s) = L_d(\bar{\mu}_d^*)u^*(s) \qquad (7)$$

$$\leqq L_d(\bar{\mu})u^*(s) \quad {}^\forall\bar{\mu}\in P(A)$$

をみたす定常多重戦略 $\bar{\mu}_d^*$ が存在する．この結果（7）より

$$u^*(s) = \sum_{t=1}^{\infty} \beta^{t-1} E_{\bar{\mu}_d^*}[\langle d, r(s_t, \bar{\mu}_d^*) \rangle \mid s_1 = s] \tag{8}$$
$$= \langle d, I(\bar{\mu}_d^*)(s) \rangle$$

を得る．一方（7）の不等式より，$\pi \in \Pi$ と time t での s_t に対して

$$u^*(s_t) \leqq L_d(\pi) u^*(s_t) \tag{9}$$
$$= \langle d, r(s_t, t, \pi) \rangle + \beta \sum_{s_{t+1}} u^*(s_{t+1}) q(s_{t+1} \mid s_t, t, \pi)$$

を得る．よって（9）より u^* への代入をくり返して

$$u^*(s) \leqq \sum_{t=1}^{\infty} \beta^{t-1} E_{\pi}[\langle d, r(s_t, t, \pi) \rangle \mid s_1 = s]$$
$$= \langle d, I(\pi)(s) \rangle \tag{10}$$

が得られる．（8）と（10）より

$$\langle d, I(\bar{\mu}_d^*)(s) \rangle \leqq \langle d, I(\pi)(s) \rangle \qquad {}^\forall \pi \in \Pi .$$

したがって，補助定理１より

$$I(\bar{\mu}_d^*)(s) \in Ext[E_s \mid D],$$

これは定理１の結果を証明している．

次に定理１の逆を得るために凸錐 $D = L \cup \{0\}$ に関する部分集合 E の凸性を導入する必要がある．

__定義２.__ $R^m \supset E$ に対して $E + D$ が R^m の中で凸集合ならば，集合 E は D-凸（D-convex）と呼ぶ．

定理2.　多重戦略 π^* が初期状態 s に対する D-solution で，E_s が D-凸 で

$$cl[E_s + D - I(\pi^*)(s)] \cap (-clD) = \{0\} \tag{11}$$

が成立しているとする．このとき

$$\langle d, I(\pi^*)(s)\rangle = \min_{\pi} \langle d, I(\pi)(s)\rangle$$

をみたす $d \in L_1^+$ が存在する．

証明　$\langle d, I(\pi^*)(s)\rangle \leqq \langle d, I(\pi)(s)\rangle \quad {}^\forall \pi \in \Pi$

をみたす $d \in L^+$ が存在しないと仮定すると，

$$(E_s + D - I(\pi^*)(s))^* \cap L^+ = \phi$$

が成立する．L^+ は凸集合であるから　2つの凸集合に対する分離定理より

$$\langle \hat{d}, v\rangle \geqq \langle \hat{d}, v'\rangle \quad {}^\forall v \in (E_s + D - I(\pi^*)(s))^*,\ {}^\forall v' \in L^+ \tag{12}$$

をみたす $\hat{d} \neq 0$ が存在する．このとき $0 \in (E_s + D - I(\pi^*)(s))^*$ より

$$\hat{d} \in -(L^+)^* = -clD. \tag{13}$$

また(12)式で $v' \to 0$ とすることによって

$$\langle \hat{d}, v\rangle \geqq 0 \quad {}^\forall v \in (E_s + D - I(\pi^*)(s))^*.$$

E_s は D-凸 より，$[E_s + D - I(\pi^*)(s)]$ は凸錐であるから

$$\hat{d} \in (E_s + D - I(\pi^*)(s))^{**} = cl[E_s + D - I(\pi^*)(s)]. \tag{14}$$

よって

$$\hat{d} \in cl[E_s + D - I(\pi^*)(s)] \cap (-clD)$$

となり　仮定(11)に反する, i.e.,

$$\langle \hat{d}, I(\pi^*)(s) \rangle \leqq \langle \hat{d}, I(\pi)(s)+x \rangle \quad {}^\forall x \in D,\ {}^\forall \pi \in \Pi .$$

上の式で　$x=0$　とすることより

$$\langle \hat{d}, I(\pi^*)(s) \rangle \leqq \langle \hat{d}, I(\pi)(s) \rangle \quad {}^\forall \pi \in \Pi . \tag{15}$$

$L \ni e=(1,1,\cdots,1)$ より $\langle \hat{d}, e \rangle > 0$ であるから，これで(15)の両辺を割り　$d = \hat{d}/\langle \hat{d}, e \rangle$　とおくことによって

$$d \in L_1^+$$

で

$$\langle d, I(\pi^*)(s) \rangle \leqq \langle d, I(\pi)(s) \rangle \quad {}^\forall \pi \in \Pi$$

が成立する．よって定理は証明された．

<u>定義 3.</u>　　$\langle d, I(\pi^*)(s) \rangle = \min_{\pi} \langle d, I(\pi)(s) \rangle$
をみたす割引率 $d \in L_1^+$ を初期状態 s に対する π^* の D-multiplier と呼ぶ．

<u>定理 3.</u>　　d_1 と d_2 が初期状態 s に対する π_1^* と π_2^* の D-multiplier とすると，

$$\langle d_1 - d_2,\ I(\pi_1^*)(s) - I(\pi_2^*)(s) \rangle \leqq 0 \tag{16}$$

が成立する．

証明　　D-multiplier の定義より

$$\langle d_1, I(\pi_1^*)(s) \rangle \leqq \langle d_1, I(\pi_2^*)(s) \rangle$$

と

$$\langle d_2, I(\pi_2^*)(s) \rangle \leqq \langle d_2, I(\pi_1^*)(s) \rangle$$

が成立する．よってこの２つの式を加えて（16）が得られる．

§4. D-solution と super-gradient について

$L_0^+ \equiv L^+ \cup \{0\}$ の上に初期状態 s に対する lower support function K_s を次のように定義する．

$$K_s(d) \equiv \inf_{\pi} \langle d, I(\pi)(s) \rangle \quad \forall d \in L_0^+.$$

<u>定義4.</u>　$f : L_0^+ \subset R^m \to R$ について

$$f(d) - f(\hat{d}) \leqq \langle d - \hat{d}, x \rangle \quad \forall d \in L_0^+ \qquad (17)$$

をみたす $R^m \ni x$ が存在すれば，この f は $\hat{d}$ で super-differentiable と呼ぶ．さらに，このようなベクトル x は $\hat{d}$ での f の super-gradient と呼び，super-gradient の全体は $\hat{\partial} f(\hat{d})$ と書かれ $\hat{d}$ での f の super-differential と呼ぶ．

<u>定理4.</u>　多重戦略 π^* が初期状態 s に対する D-multiplier $\bar{d} \in L_1^+$ に関する D-solution であるための必要十分条件は $I(\pi^*)(s)$ が $L_1^+ \ni \bar{d}$ で lower support function K_s の super-gradient であることである，i.e.,

$$I(\overline{\pi}^*)(s) \in \tilde{\partial} K_s(\overline{d}) .$$

証明　super-gradient の定義より（17）は次のように書ける.

$$K_s(d) - K_s(\overline{d}) \leqq \langle d - \overline{d}, I(\overline{\pi}^*)(s) \rangle \quad {}^\forall d \in L_0^+ .$$

上の式で $d=0$ とおくと

$$K_s(\overline{d}) \geqq \langle \overline{d}, I(\overline{\pi}^*)(s) \rangle ,$$

i.e.,

$$\langle \overline{d}, I(\overline{\pi}^*)(s) \rangle \leqq \langle \overline{d}, I(\overline{\pi})(s) \rangle \quad {}^\forall \overline{\pi} \in \Pi . \tag{18}$$

（18）式に　補助定理１を適用して

$$I(\overline{\pi}^*)(s) \in Ext[E_s | D]$$

が得られ，$\overline{\pi}^*$ は初期状態 s に対する D-solution となる.

逆に $\overline{\pi}^*$ が初期状態 s に対する D-multiplier $\overline{d} \in L_1^+$ に関連している D-solution とすると

$$K_s(\overline{d}) = \langle \overline{d}, I(\overline{\pi}^*)(s) \rangle . \tag{19}$$

さらに

$$K_s(\overline{d}) \leqq \langle d, I(\overline{\pi}^*)(s) \rangle , {}^\forall d \in L_0^+ \tag{20}$$

であるから，（19）式から（20）式を引くことによって

$$K_s(d) - K_s(\overline{d}) \leqq \langle d - \overline{d}, I(\overline{\pi}^*)(s) \rangle$$

が得られる．よって

$$I(\overline{\pi}^*)(s) \in \tilde{\partial} K_s(\overline{d}) .$$

かくて定理は証明された.

References

1. J.P.Aubin, Mathematical Methods of Game and Economic Theory, North-Holland Publishing Company, 1979.
2. J.P.Aubin, Applied Functional Analysis, Wiley-Interscience, New York, 1979.
3. K.Fan, Fixed point and minimax theorem in locally convex topological linear spaces, Proc. Nat. Acad. Sci., U.S.A. 38 (1952), 121 - 126.
4. K.Fan, Minimax theorems, Proc. Nat. Acad. Sci., U.S.A. 39 (1953), 42 - 47.
5. H.C.Lai and K.Tanaka, Noncooperative n-person game with a stopped set, J. Math. Anal. Appl., 88 (1982), 153 - 171.
6. H.C.Lai and K.Tanaka, A noncooperative n-person semi-Markov game with a separable metric space, Applied Math. and Opti., 11 (1984), 23 - 42.
7. H.C.Lai and K.Tanaka, On an N-person noncooperative Markov game with a metric state space, J. Math. Anal. Appl., 101 (1984), 78 - 96.
8. P.L.Yu, Cone convexity, cone extreme points, and nondominated solutions in decision problems with multiobjectives, Journal of Optimization Theory and Applications, 14 (1974), 319 - 377.